Functional Linear Models for Scalar Responses

Pang Du

Department of Statistics
Virginia Tech

FLMs with scalar responses

- With functional responses and multivariate independent variables we could estimate the regression coefficient functions without the necessity of using roughness penalties.
- The same for the concurrent model with functional responses and functional independent variables.
- Now we look at a scalar response predicted by a functional independent variable, and a *roughness penalty* or *regularization* is indispensable.

A model for total annual precipitation

- $y_i = \text{LogPrec}_i$: the logarithm of total annual precipitation at weather station i.
- Our model:

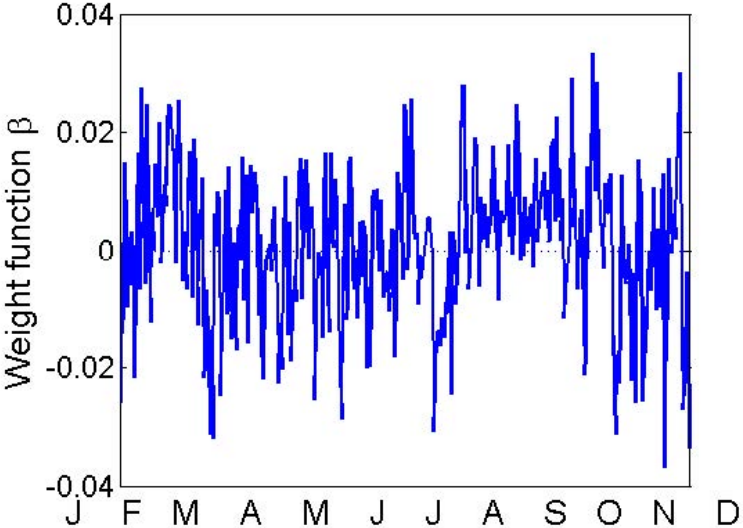
$$\text{LogPrec}_i = \alpha + \int_0^T \text{Temp}_i(s)\beta(s)ds + \epsilon_i.$$

- We can think of the function values $\text{Temp}(s)$ associated with each fixed s as a separate scalar independent variable.
- If so, we have enough fitting power at our disposal to fit any number of responses, and certainly only 35 of them.

A bad idea

- If we use the discrete daily temperature averages, we have 365 plus 1 for constant α independent variables to fit 35 responses.
- Using the Moore-Penrose generalized inverse to keep us out of trouble, we get the following estimate of β.

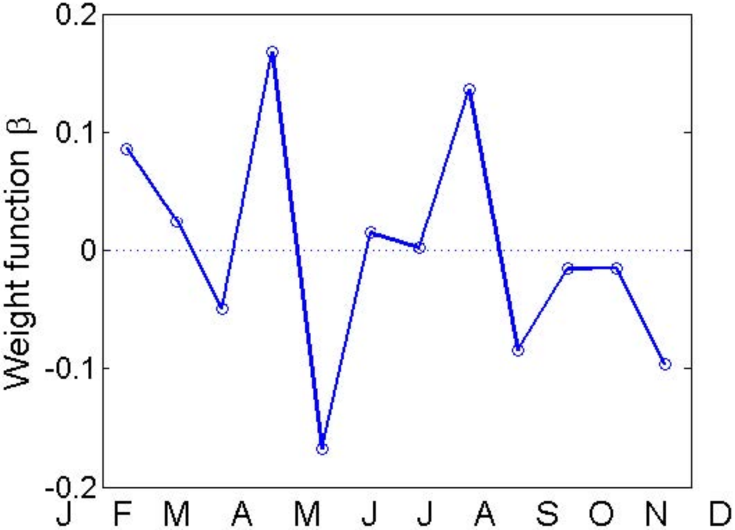
Log precipitation: a bad fit



A better idea

- If we use the 12 monthly temperature averages for each station, we have 12 plus 1 for constant α independent variables to fit 35 responses.
- A simple multiple linear regression can do the job.
- $R^2 = 0.84$: a rather successful fit.
  F-ratio is 9.8 with 12 and 22 dfs: significant at 0.01 level.
  Std. err. is 0.34, std. dev. of response is 0.69.

Log precipitation: a better fit



What do we see?

- The temperatures in the months of January, April, May, August, September, and December are important.
- A functional approach can achieve better interpretation.

Basis expansions

- Expand β into K_β Fourier basis functions:

$$\beta(s) = \sum_{k=1}^{K_\beta} b_k \theta_k(s) = \boldsymbol{\theta}'\mathbf{b}.$$

- Expand covariate functions Temp_i into K_z Fourier basis functions:

$$\text{Temp}_i(s) = \sum_{k=1}^{K_z} c_{ik} \psi_k(s) = \boldsymbol{\psi}'\mathbf{c}_i.$$

  Let $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_n)'$ be the coefficient matrix of **Temp**.

- Both K_β and K_z are chosen to be small numbers (certainly smaller than $N = 35$).

Rewrite the model

- Define the K_z by K_β matrix $\mathbf{J}_{\psi\theta} = \int \boldsymbol{\psi}(s)\boldsymbol{\theta}'(s)ds$.
- The model in vector-matrix form:

$$\mathbf{y} = \alpha + \mathbf{C}\mathbf{J}_{\psi\theta}\mathbf{b} + \boldsymbol{\epsilon}.$$

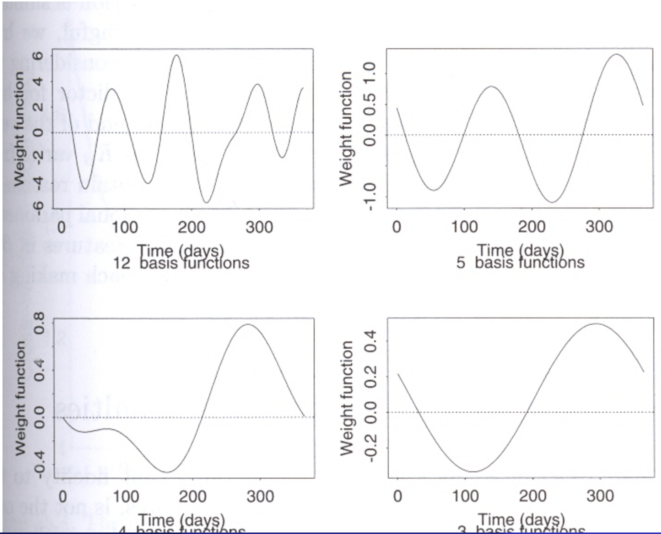
- Define $\boldsymbol{\zeta} = (\alpha, b_1, \ldots, b_{K_\beta})'$ and $\mathbf{Z} = [\mathbf{1} \quad \mathbf{C}\mathbf{J}_{\psi\theta}]$.
- This reduces to a standard multiple linear regression problem and the estimate is:

$$\hat{\boldsymbol{\zeta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}.$$

Our choices of K_β and K_z

- $K_z = 12$.
- Preferably $K_\beta \leq K_z$ (why?).
- Tried $K_\beta = 12, 5, 4, 3$.

Log precipitation: finite basis expansion



What do we see?

- $K_\beta = 12$: similar to monthly average discretization, not very informative.
- $K_\beta = 5$: more oscillations than the fits from even smaller K_β. Are all the oscillations true trends?
- $K_\beta = 4, 3$: show similar trends suggesting that a predictor for high annual precipitation is a relatively high temperature towards the end of the year.
- Drawback: the model complexity appears to have a discrete jump from $K_\beta = 4$ to $K_\beta = 5$.
- The alternative roughness penalty approach can provide a smooth change in model complexity with a varying smoothing parameter.

The penalized least squares

- The basis expansion approach has discrete jumps in model complexity.
- Using a roughness penalty gives us continuous control over smoothness and other advantages.
- Here is the penalized least squares criterion:

$$\mathrm{PENSSE}_\lambda(\alpha,\beta) = \sum_{i=1}^{N}[y_i - \alpha - \int z_i(s)\beta(s)ds]^2 + \lambda \int [L\beta(s)]^2 ds.$$

Choosing a roughness penalty

- Let's penalize **harmonic acceleration** because we want $\beta(s)$ to be periodic:

$$L\beta(s) = \left(\frac{2\pi}{365}\right)^2 D\beta(s) + D^3\beta(s).$$

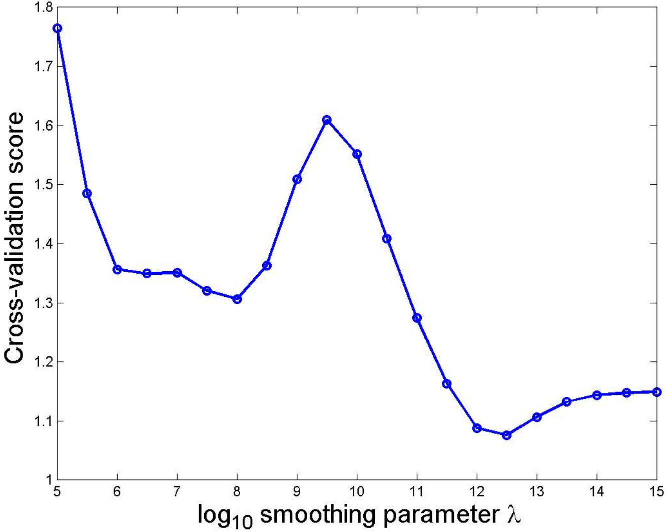
- We choose the smoothing parameter λ by minimizing the cross-validation criterion.
- Let $\alpha_\lambda^{(-i)}$ and $\beta_\lambda^{(-i)}$ be the estimates using all the responses except for y_i.
- The criterion to be minimized is

$$\mathrm{CV}(\lambda) = \sum_{i=1}^{N}[y_i - \alpha_\lambda^{(-i)} - \int z_i(s)\beta_\lambda^{(-i)}(s)ds]^2.$$

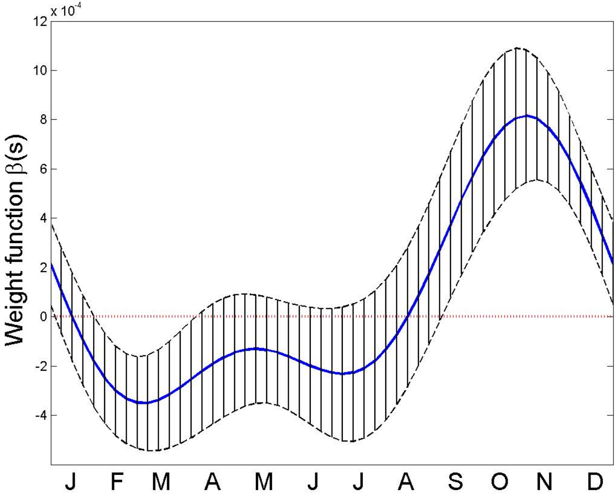
More details

- We use 65 Fourier basis functions to represent the temperature curves.
- We use 35 Fourier basis functions to represent β.
- The final fit uses $\lambda = 10^{12.5}$.

A plot of CV(λ)



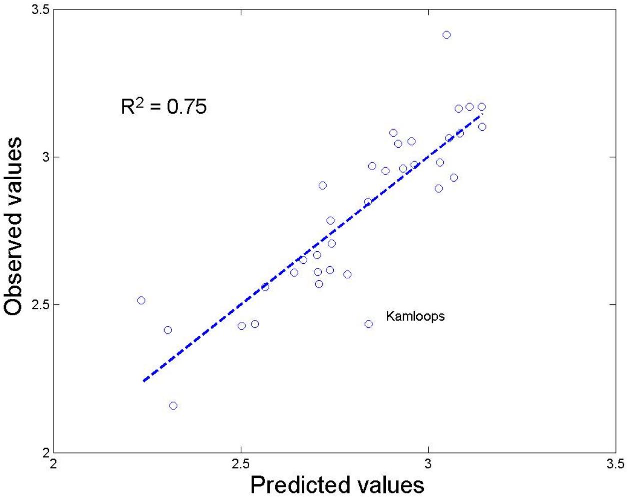
Log precipitation: fit by roughness penalty



What do we see?

- CLs in the earlier summer months contain 0: the influence of temperature in that period on the annual precipitation is not important.
- A strong peak in the late fall followed by a valley in the early spring
    - there is a contrast between fall and early temperature influence with more emphasis on the autumn.
    - this pattern favors weather stations that are comparatively warm in October and cool in spring, and where spring comes early.
    - consistent with previous analysis results for the Pacific and Atlantic stations with marine climiates, where the seasons are later than average and the fall weather is warm relative to the inland stations.

Log precipitation: Observed vs. prediction



What do we see?

From the plot of observed versus predicted values,

- $R^2 = 0.75$ and most points are close to the diagonal line: a good fit.
- Kamloops: observed value is 2.5 and predicted value is 2.9, a bit off.
    - Kamloops is deep in the Thompson River Valley where the rain clouds usually just pass on by.
    - Can be detected by diagnostic plot to be introduced.

Computation: estimation

- Expand covariates z_i into K_z basis functions ψ_m and the regression function β into K_β basis functions θ_k.
- Define the penalty matrix $\mathbf{R} = \int [L\boldsymbol{\theta}(s)][L\boldsymbol{\theta}(s)]' ds$,
- the K_z by K_β matrix $\mathbf{J}_{\psi\theta} = \int \boldsymbol{\psi}(s)\boldsymbol{\theta}'(s) ds$,
- and the augmented vector $\boldsymbol{\zeta} = (\alpha, b_1, \ldots, b_{K_\beta})'$.
- Define $\mathbf{Z} = [\mathbf{1} \quad \mathbf{CJ}_{\psi\theta}]$ and augment $\mathbf{R}$ to $\mathbf{R}_0$ to include a leading column and row of $K_\beta + 1$ zeroes.
- This reduces to a familiar ridge regression problem and the estimate is:

$$\hat{\zeta} = (\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R}_0)^{-1} \mathbf{Z}'\mathbf{y}.$$

Computation: confidence limits

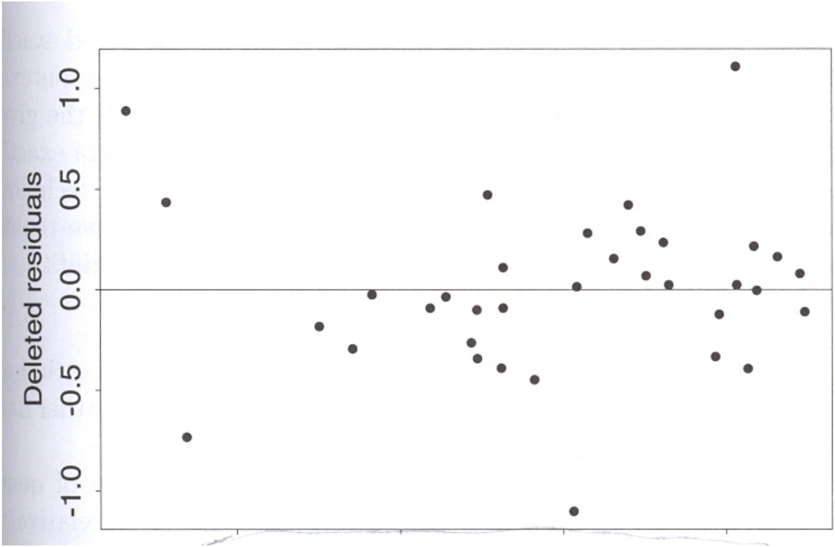
- The variance-covariance matrix Σ_e in the previous functional linear models now reduces to a scalar σ_e^2.
- The sampling variance of $\hat{\zeta}$ is:

$$\text{Var}(\hat{\zeta}) = \hat{\sigma}_e^2(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{R}_0)^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{R}_0)^{-1}.$$

Computation: “Hat” matrix

- Let
$$\mathbf{S} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{R}_0)^{-1}\mathbf{Z}'$$
be the **hat matrix** of the smoothing procedure.
- Some calculation shows that
$$\mathrm{CV}(\lambda) = \sum_{i=1}^{N} \left(\frac{y - \hat{y}_i}{1 - S_{ii}} \right)^2,$$
where S_{ii} is the ith diagonal element of $\mathbf{S}$.
- *Effective degrees of freedom* can be defined as trace($\mathbf{S}$) (=4.7 in our precipitation example).
- *Deleted residuals*: $(y_i - \hat{y}_i)/(1 - S_{ii})$.

Log precipitation: deleted residuals



FLMs with scalar responses

- Either dimension reduction or regularization is essential when the dimensionality of the covariate exceeds the dimensionality of the response.
- Having gone through the most development since the publication of the book (may cover some recent developments in later classes).